DOCUMENT RESUME

ED 435 087 CS 013 675

AUTHOR Kranzler, John H.; Miller, M. David
TITLE An Examination of Racial/Ethnic and Gender Bias on Curriculum-Based Measurement of Reading.
PUB DATE 1999-04-00
NOTE 18p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Blacks; Curriculum Based Assessment; Elementary Education; Racial Bias; *Reading Achievement; Reading Comprehension; Reading Research; Sex Bias; *Test Bias; Test Validity; Whites
IDENTIFIERS African Americans

ABSTRACT

A study examined racial/ethnic and gender bias on curriculum-based measurement (CBM) of reading with African-American and Caucasian male and female regular education students across grades 2-5. Simultaneous multiple regression analyses were conducted by grade to examine group differences on CBM as an estimate of reading comprehension. Regression equations were estimated with CBM, gender, race/ethnicity, and the interactions of gender and race/ethnicity with CBM. Results indicated that CBM is a biased indicator of reading comprehension. Although no evidence of bias was found at the second and third grades, intercept bias was found for racial/ethnic groups at the fourth and fifth grades, and intercept and slope bias were found for gender at the fifth grade. Implications suggest that the meaning of CBM scores differs across race/ethnicity or gender, or both, at certain grade levels. CBM performance over-estimates the reading comprehension of African American students and under-estimates that of Caucasians; and at grade 5, CBM performance overestimates the reading comprehension of girls and under-estimates that of boys. Mean differences between boys and girls were also much greater at lower levels of CBM performance than at higher levels. These findings raise issues concerning the use of CBM as a screening measure and in determining eligibility for and termination of special education and related services. (Author/RS)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 087

An Examination of Racial/Ethnic and Gender Bias on Curriculum-Based Measurement of Reading

John H. Kranzler, M. David Miller,
& LuAnn Jordan
University of Florida

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J Kranzler
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CS 013 675

Abstract

We examined racial/ethnic and gender bias on curriculum-based measurement (CBM) of reading with African-American and Caucasian male and female regular education students across Grades 2-5. Simultaneous multiple regression analyses were conducted by grade to examine group differences on CBM as an estimate of reading comprehension. Regression equations were estimated with CBM, gender, race/ethnicity, and the interactions of gender and race/ethnicity with CBM. Results of this study indicated that CBM is a biased indicator of reading comprehension. Although no evidence of bias was found at the second and third grades, intercept bias was found for racial/ethnic groups at the fourth and fifth grades, and intercept and slope bias were found for gender at the fifth grade. Implications of these results for the use of CBM with different groups are potentially important, because they suggest that the meaning of CBM scores differs across race/ethnicity or gender, or both, at certain grade levels. According to our findings, at Grades 4 and 5, CBM performance over-estimates the reading comprehension of African American students and under-estimates that of Caucasians. Our results also suggest that, at Grade 5, CBM performance over-estimates the reading comprehension of girls and under-estimates that of boys. Mean differences between boys and girls were also much greater at lower levels of CBM performance than at higher levels. These findings raise issues concerning the use of CBM as a screening measure and in determining eligibility for and termination of special education and related services.

Introduction

Forness, Kavale, Blum, and Lloyd (1997) recently summarized the results of 18 meta-analyses of the effectiveness of interventions in special education and related services. According to their results, one of the most effective strategies involves the integration of formative assessment of academic performance and positive reinforcement of effort and accomplishment. Formative assessment was conducted in these studies with curriculum-based measurement (CBM).

Deno (1985, 1989) developed CBM to inform the instructional decision-making of special education teachers. CBM refers to a specific set of brief, fluency-based measures of basic academic skills (viz., reading, math, writing, and spelling). More recently, proponents of CBM have argued for its use as a screening measure and in determining eligibility for special education and related services (see Shinn, 1998).

The role of CBM within a comprehensive model of academic problem-solving is outlined in Table 1. As shown in this table, "CBM-guided decision making relies primarily on a norm-referenced approach" (Shinn & Habedank, 1992, p. 12). Given its reliance on norm-referenced interpretation, an important assumption underlying the CBM problem-solving model is that scores have the same meaning for all children at a particular grade level in a particular locale. Moreover, despite empirical support for its use in regular and special education, the validity of CBM reading with children and youth from diverse linguistic and racial/ethnic backgrounds has yet to be thoroughly examined. Only one previous study has examined the issue of test bias on CBM with African American students (i.e., Knoff & Dean, 1994). Unfortunately, results of this study are inconclusive due to use of an inadequate definition of bias.

What Is Test Bias?

Test bias refers to systematic measurement error or estimation related to the use of tests with two or more groups (Reynolds, 1999). "A biased test yields scores that mean something different for persons of one group than for persons of another group, even when two persons from different groups have identical scores on the test" (Jensen, 1981, p. 137). Criteria for determining test bias fall into three major categories: situational bias, internal indicators of bias, and external indicators of bias. Internal indicators of bias are most important for the practical use of tests:

> We are concerned here with a test's usefulness as a predictor of a particular criterion and with whether the test has the same predictive efficiency in different subpopulations. . . . Predictive bias means systematic error (as contrasted to random errors of measurement) in the prediction of the criterion variable for persons of different subpopulations as a result of basing prediction on a common regression equation for all persons regardless of their subpopulation membership, or basing prediction for persons of one subpopulation on the regression equation derived on a different subpopulation. (Jensen, 1980, p. 380)

Examination of external indicators of bias, or predictive bias, as it is often called, is not limited to situations involving the prediction of a criterion at some distant point in the future. This category of test bias also encompasses situations in which there is a short interval between the test and criterion measurements or no interval at all.

Aims of This Study

Examination of external indicators of bias are the most germane for the practical use of CBM, because: (a) Deno's (1989) problem-solving model depends on norm-referenced interpretation of CBM performance as an estimate of current scholastic achievement for screening and for determining eligibility for and termination of special education and related services; and (b) CBM has been proposed as a substitute for more time-consuming and expensive ways of measuring basic academic skills, such as nationally standardized tests of scholastic achievement (e.g., see Shinn, 1989).

It is important to note that in the CBM problem-solving model academic expectations are based on the performance of "typical" same-grade peers, without regard to subpopulation membership. A common set of norms is used for all students at a particular grade level. Despite the fact that CBM scores are "referenced to the performance of a local norm group that is presumably maximally similar in acculturation (e.g., learning opportunities, background experiences) to the student in question" (Shinn, Nolet, & Knutson, 1990, p. 292), use of local norms does not guarantee that CBM is equally valid and unbiased for all groups of students. The interpretation of CBM scores might not be biased in favor of or against certain subpopulations; but it might be. The only way to determine the presence or absence of test bias is by analyzing empirical data from two or more groups with objective statistical methods (see Jensen, 1980; Reynolds, 1995).

The aim of this study was to examine racial/ethnic and gender bias on CBM reading as an index of reading comprehension with African-American and Caucasian male and female regular education students across Grades 2-5.

Method

Participants

Participants in this study were 326 students (170 boys, 156 girls) in Grades 2 to 5 (ns = 84, 76, 94, and 72, respectively), selected randomly from the general education classes of a public elementary school in North Central Florida. None of the participants was receiving special education services. In terms of racial/ethnic group composition, the sample consisted of 225 Caucasians and 79 African Americans. The primary language of all participants was English. Table 2 shows the number and percentage of boys and girls and of African American and Caucasian students in the sample across grade level. All participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 1992).

Procedures

Participants were administered six curriculum-based measures of reading fluency in one test session in March as part of a school-wide CBM validity study. Generalizability[3] of CBM in this study exceeded .90 for all grade levels (Miller & Jordan, 1996). Trained graduate students administered the CBM probes. The standardized, norm-referenced test of reading comprehension was administered in the spring under standardized conditions.

Instruments

Curriculum-Based Measurement of Reading. Administration and scoring of the curriculum-based measures of reading fluency followed standardized procedures. The CBM probes were chosen from the reading textbooks used in Grades 2-5 in the local school district (viz., Ginn Basal Readers). Passages of 250 words or more were randomly selected. Passages consisting of prose, plays, and poetry were eliminated, as well as stories with a high degree of dialogue. From each story chosen, a Frye (1968) readability index was calculated on passages of 250 words. Differences in appearance between probes were controlled by retyping selected passages in a font and type size similar to the Ginn Basal Readers. One form of each passage was created for students and one for examiners. Participants read aloud from passages selected at random for one minute, while the examiner recorded the number of words read correctly. The generalizability coefficients for the CBM probes used in this study exceeded .90 at each grade level. Due to the considerable consistency of scores across CBM probes, the mean of the six probes was used in all analyses as the measure of CBM reading.

California Achievement Test (CAT). The California Achievement Test (CAT) is a major standardized achievement test battery covering reading, writing, mathematics, science and social studies from Grades K to 12. According to reviews in The Tenth Mental Measurements Yearbook by Airasian (1989) and Waldrop (1989), the CAT, Forms E and F, have high internal consistency estimates and high content validity. The scaled scores reported for the CAT were equated through a 3-parameter logistic model (IRT). Overall, the CAT provides good psychometric data pertaining to content validity, although construct validity is not addressed adequately (as is typical of most achievement test batteries).

Statistical Analyses

- Descriptive statistics were calculated for each racial/ethnic group and gender across grade.
- Pearson product-moment correlations were used to examine the relationship between CBM reading and CAT Reading Comprehension.
- Mean differences between racial/ethnic and gender groups were examined with t-tests.
- Simultaneous multiple regression was used to examine the presence or absence of racial/ethnic and gender bias on CBM reading as an estimate of CAT reading comprehension. At each grade level, a multiple regression equation was estimated with CBM, gender, race/ethnicity, and the interactions of gender and race/ethnicity with CBM reading.
- All analyses were conducted by grade level, because the CBM probes and CAT items differed across grade. Because the passages for the CBM probes are linked to the curriculum at each grade level, comparisons across grade are inappropriate.
- A biased test was defined as one in which the regression lines of the groups differed significantly in slopes ($b_{y.x}$) or intercepts (k). An unbiased test was defined as one in which the regression lines of the two groups (i.e., $b_{y.x}$ or k) did not differ significantly. In these analyses, the effects of gender and race/ethnicity addressed the issue of intercept bias; whereas the interactions of gender and race/ethnicity addressed slope bias.

Discussion

- Results of simultaneous multiple regression analyses indicated that CBM fails as an unbiased indicator of current reading comprehension.
- Although no evidence of bias was found at the second and third grades, intercept bias was found for racial/ethnic groups at the fourth and fifth grades, and intercept and slope bias were found for gender at the fifth grade.
- Because CBM reading is not an unbiased test, the meaning of scores on CBM differed across race/ethnicity and gender at particular grade levels in this study.
- At Grades 4 and 5, CBM performance over-estimated the reading comprehension of African American students and under-estimated that of Caucasians. In addition, at Grade 5, CBM performance over-estimated the reading comprehension of girls and under-estimated that of boys, although differences between boys and girls on CBM reading were much greater at lower levels of performance than they were at higher levels.

Implications

- In the CBM problem-solving model, placement decisions are based on norm-referenced interpretation of CBM performance. Underlying this model is the assumption that the same score on CBM is interpreted to reflect the same level of current academic achievement for all groups.
- If CBM reading is biased, however, systematic error may exist in the estimation of reading comprehension for children of different groups at certain grade levels, when estimates are based on a common set of expectations for all students without regard to their racial/ethnic or gender group membership.
- The impact of bias in estimation will be greatest for students whose CBM performance falls near the cutting score that is used for eligibility determination.
- In this study, for African American students at Grades 4 and 5, and for girls at Grade 5, systematic over-estimation of reading comprehension will result in the under-identification of children whose reading comprehension is in need of remediation, as defined by the CBM problem-solving model.
- Evidence of bias does not mean that CBM should be rejected outright or that it should be used only with certain groups, however. Systematic under- and over-identification can be eliminated by using different estimates of performance and different cut-off scores across groups for screening and for determining eligibility for and termination of special education and related services.

Conclusion

According to the results of Forness et al.'s (1997) recent review of meta-analyses on the effectiveness of interventions in special education and related services, one of the most effective strategies involves the integration of formative assessment of academic skills—that is, CBM—with positive reinforcement of effort and accomplishment. Notwithstanding this impressive finding, the broader use of CBM in a comprehensive model of problem-solving (see Deno, 1989), in which CBM data are not only used for monitoring progress in the curriculum, but also for screening and for determining eligibility for and termination of special education and related services, will depend, at least in part, upon the results of further research on CBM test bias.

References

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Deno, S. L. (1989). Curriculum-based measurement and alternative special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), Curriculum-based measurement: Assessing special children (pp. 1-17). New York: Guilford.

Jensen, A. R. (1981). Straight talk about mental tests. New York: The Free Press.

Knoff, H. M., & Dean, K. R. (1994). Curriculum-based measurement of at-risk students' reading skills: A preliminary investigation of bias. Psychological Reports, 75, 1355-1360.

Reynolds, C. R. (1995). Test bias and the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), International handbook of personality and intelligence (pp. 545-573). New York: Plenum.

Shinn, M. R. (1989). Curriculum-based measurement: Assessing special children. New York: Guilford.

Shinn, M. R., Nolet, V., & Knutson, N. (1990). Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology (pp. 287-308). Washington, DC: National Association of School Psychologists.

Author Notes

1. In Deno's (1989) problem-solving model, discrepancies between current and expected CBM performance warranting further assessment are defined as either

(a) CBM scores that fall below the 10th percentile in comparison to same-grade peers; or
(b) CBM scores that are half that of typical peers at the same-grade level.

2. Mean differences between racial/ethnic groups have declined in recent years, however. For example, on the Scholastic Aptitude Test (SAT) the mean difference between Caucasians and African Americans (in standard deviation units) decreased from 1.16 to 0.88 on the SAT-Verbal and from 1.27 to 0.92 on the SAT-Math between 1976-1993 (Herrnstein & Murray, 1994). Reductions in mean group differences have also been documented on at least some IQ tests (e.g., Lynn, 1996).

Table 1

CBM Problem-solving Model Decisions, Measurement and Evaluation Activities, and Specific Tasks

Problem-Solving Decision Step	Measurement Activities	Evaluation Activities	Specific Tasks
I. Problem Identification	Observe and record student performance	Decide whether a discrepancy exists between actual and expected performance	Peer-referenced assessment (Screening)
II. Problem Certification	Describe differences between actual and expected performance in the context of likelihood of general education resources	Decide whether discrepancies are important enough to require special services for problem resolution	Conduct survey-level assessment, evaluate general educational modifications (Eligibility Determination)

Table 1 (continued)

III.	Exploring Alternative	: Determine probable performance Select the program reform (i.e., Write long-term goal(s), Solutions (IEP goal setting; improvements (goals) and costs intervention) to be tested determine curriculum level Intervention planning) associated with different and necessary pre-skills interventions required for success
IV.	Evaluating Solutions and	 Monitor implementation and Determine whether intervention Collect progress monitoring Making Modifications change in student performance is effective or should be modified data and compare with IEP (Progress Monitoring) goals.

Table 1 (continued)

V. Problem Solution

Observe and record student
Decide whether discrepancies
Repeat peer-referenced
(Program Termination)
differences, if any, between actual
discrepancies are
significant. If
assessment
and expected performance
not, program may be terminated

Note. Adapted from Shinn and Habedank (1992).

Table 2

Frequencies by Gender and Racial/Ethnic Group Across Grade

Grade		2	3	4	5
Gender					
Girls	N	39	40	43	34
	%	46.43	52.63	45.74	47.22
Boys	N	45	36	51	38
	%	53.57	47.37	54.26	52.78
Total	N	84	76	94	72
Racial/Ethnic Group					
African American	N	17	24	19	19
	%	22.37	35.29	20.88	27.54
Caucasian	N	59	44	72	50
	%	77.63	64.71	79.12	72.46
Total	N*	76	68	91	69

Note. The total N in this table for racial/ethnic group is less than that for gender due to the small number of students identified as Asian American or Hispanic American.